



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/610,684	07/02/2003	Scott Shepard	02-4038	2914

25537 7590 02/03/2009

VERIZON
PATENT MANAGEMENT GROUP
1320 North Court House Road
9th Floor
ARLINGTON, VA 22201-2909

EXAMINER

BAKER, MATTHEW H

ART UNIT	PAPER NUMBER
----------	--------------

2626

NOTIFICATION DATE	DELIVERY MODE
-------------------	---------------

02/03/2009

ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

patents@verizon.com

Office Action Summary

Application No.

10/610,684

Applicant(s)

SHEPARD ET AL.

Examiner

Matthew Baker

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period **will** apply and **will** expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply **will**, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 16 September 2008.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-47 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-47 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 02 July 2003 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____.
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date: _____.
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Applicant has filed an amendment received on 16 September 2008. Claims 1-47 are pending.

Applicant has amended claims 1, 20, 21, 40, and 47.

Applicant has argued to traverse the rejection of claims 1-11, 13-31, 33-38, 40, and 44-45 under 35 USC 103(a) as being unpatentable over the Foster ("Target-Text Mediated Interactive Machine Translation, 1997(in view of Schulz (US 6,360,237).

Applicant has argued to traverse the rejection of claims 41 and 46 under 35 USC 103(a) as being unpatentable over the Foster ("Target-Text Mediated Interactive Machine Translation, 1997) in view of Schulz (US 6,360,237) in further view of Saindon (US 6,820,055).

Applicant has argued to traverse the rejection of claims 12, 19, 32, 39, 42, 43, and 47 as being unpatentable over the Foster ("Target-Text Mediated Interactive Machine Translation, 1997(in view of Schulz (US 6,360,237) in further view of (Shiotani (US 4,814,988).

Applicant has argued that Foster is not combinable with Schulz.

Claim Objections

Claim 47 objected to because of the following informalities: the claim was presented as new but is referred to in arguments as amended. The claim will be treated as an amended claim. Appropriate correction is required.

Response to Arguments

2. Applicant's arguments filed 16 September 2008 with respect to the rejections under 35 USC 103 have been fully considered but they are not persuasive.
3. Applicant's argument filed with respect to the proper combination of Foster in view of Schulz has been fully considered but is not persuasive.
4. Applicant argues the Foster and Schulz do not disclose or suggest "receiving translation actually made by the user of the portion of the audio signal (Remarks, p. 16)." Applicant admits that "Foster does involve a human translator (Remarks, p.16). Applicant's explanation of Foster, pg. 179, section 3, paragraph 1 is accurate, however applicant's interpretation of the citation is not convincing. Applicant concludes that since Foster's method *could* include a machine-human combination to achieve proper translation that Foster's method does not teach the limitation of "translation actually made by the user," and that Foster's teaching of machine completion leads away from the claimed invention.

Foster's method *can* use a machine-human combination, however in using this method the user, without question, can complete 100% of the translation themselves (as opposed to the 30% cited repeatedly by applicant). The percentages cited on page 192 of Foster are meant to provide evidence of machine aid as beneficial to the method, but *not* necessary. Foster can be relied upon to teach translation "actually" made by a user.

The argument that Schulz does not meet the limitation of "receiving translation actually made by a user" is moot because Foster teaches this limitation.

5. Applicant argues that by changing the claim limitation from "receiving translation made by the user" to --receiving translation actually made by the user-- that the claim is clarified. The

word “actually,” even as defined by the Applicant's citation of Merriam Webster's Collegiate Dictionary (Tenth Edition) does not change the scope of the claim, and Foster nonetheless teaches the amended limitation.

6. Applicant argues that the limitation of “receiving translation actually made by the user of the portion of the audio signal” is not met because “*portion* of the audio signal” is not taught by Foster. Applicant's limited example the word “patent” and the possible portions of “pa” and “pat” is not convincing. A human translator is capable of taking into account context, probability of a word appearing, proper grammar, and semantic sense when translating, even when hearing only a portion of a word.

Applicant's conclusion that Foster teaches translation of only 30% of a word is an incorrect interpretation of the method taught at p. 192. Foster actually teaches that a user *may* be able to use 30% of the normal amount of keystrokes needed for translation *if* the user decides to accept a proposed completion suggested by a machine. In spite of this advantageous option, the user may actually translate 100% of the text (using 100% of the keystrokes needed to properly translate) without accepting any proposed completions.

7. The matching arguments for independent claims 20, 21, and 40 are not convincing for the same reasons above.

The claims which depend on claims 1, 20, 21, and 40 do not depend from allowable base claims, and the limitations in the dependent claims are taught as shown in the rejection below.

8. Applicant argues that Foster and Schulz are not combinable. The examiner stated in the previous rejection that “speech recognition systems are commonly used to convert speech to text, as indicated in Schulz (p. 4).” It is well known in the art that a translation system utilizing

Art Unit: 2626

speech recognition will convert received speech into text for translation, as well as speech (this is clearly taught in Schulz, for example at col. 1, ll. 27-34).

Applicant provides no argument as to why the references are not combinable besides general disagreement that the examiner's rationale is not satisfactory. The examiner has established *prima facie* of obviousness for motivation for combination in the previous rejection (p. 4-5), which will be repeated in the art rejection below.

In response to applicant's argument that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

In response to applicant's argument that Franz is nonanalogous art, it has been held that a prior art reference must either be in the field of applicant's endeavor or, if not, then be reasonably pertinent to the particular problem with which the applicant was concerned, in order to be relied upon as a basis for rejection of the claimed invention. See *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992).

9. Applicant argues that Shiotani and Schulz are not combinable. The examiner has established *prima facie* of obviousness for motivation for combination in the previous rejection (p. 21), which will be repeated in the art rejection below.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-11, 13-31, 33-38, 40, and 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Foster*** ("Target-Text Mediated Interactive Machine Translation" Machine Translation, 1997) in view of ***Schulz*** (6,360,237).

10. As per claims 1 and 20, ***Foster*** discloses a method and system for facilitating translation of an audio signal that includes speech to another language, comprising:

Retrieving a textual representation (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been retrieved*);

Presenting the textual representation to a user (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been presented to the user*);

Receiving selection of a segment of the textual representation for translation (page 179, section 3, first paragraph, *the translator selects a portion of the source text, usually a sentence, for translation*);

Receiving translation actually made by the user (page 179, section 3, first paragraph, *the translator selects a portion of the source text, usually a sentence, and types in the translation*).

Foster does not disclose retrieving a textual representation of an audio signal, obtaining a portion of the audio signal corresponding to the segment of the textual representation, providing the segment of the textual representation and the portion of the audio signal to the user, and receiving a translation made by the user of the portion of the audio signal. Rather, as noted above, **Foster** discloses human translation of text, without providing specifics as to where the text came from. However, speech recognition systems are commonly used to convert speech to text, as indicated in **Schulz** (column 1 lines 27-34, *speech recognition is used for transcription*). **Schulz** also discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on a screen with a specific spoken word during playback of an audio file. All of the elements of claims 1 and 20 are known in references **Foster** and **Schulz**, the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use known methods to retrieve a textual representation of an audio signal for translation in **Foster**, since it would provide automatic transcription, saving transcription costs (Schulz, column 1 lines 27-34), while enabling a user to provide fast and accurate translation of speech data.

It would also have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with **Foster**, since

Art Unit: 2626

the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

11. As per claim 21, **Foster** discloses a translation system, comprising:

Obtaining a textual representation (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been retrieved*);

Presenting the transcription to a user (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been retrieved*);

Receiving selection of a portion of the transcription for translation (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been retrieved*);

Receive from the user a translation actually made by the user of the portion of the audio signal (page 179, section 3, first paragraph, *the translator selects a portion of the source text, usually a sentence, and types in the translation*).

Foster does not disclose a memory configured to store instructions, and a processor configured to execute the instructions in memory to perform the aforementioned steps as well as obtain a transcription of an audio signal that includes speech, retrieve a portion of the audio signal corresponding to the portion of the transcription, and provide the portion of the

Art Unit: 2626

transcription and the portion of the audio signal to the user. However, **Foster** discloses a system for Interactive Machine Translation, where the user provides a translation of the source data using a machine translation system as a resource. The use of the MT system suggests the use of a computer, including memory and a processor configured to execute instructions from memory. Additionally, speech recognition systems are commonly used to convert speech to text, as indicated in **Schulz** (column 1 lines 27-34, *speech recognition is used for transcription*). **Schulz** also discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on a screen with a specific spoken word during playback of an audio file. All of the elements of claim 21 are known in references **Foster** and **Schulz**, the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a memory and processor configured to execute the instructions stored in memory in **Foster**, since a computer system can perform calculations and execute instructions extremely quickly, thus decreasing processing time and enabling a real-time application.

It would also have been obvious to one of ordinary skill in the art at the time of the invention to use known methods to retrieve a textual representation of an audio signal for translation in **Foster**, since it would provide automatic transcription, saving transcription costs (Schulz, column 1 lines 27-34), while enabling a user to provide fast and accurate translation of speech data.

It would also have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with *Foster*, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in *Schulz* (column 5 lines 55-58).

12. As per claims 2 and 22, *Foster* in view of *Schulz* disclose the method and system of claims 1 and 21, but *Foster* does not explicitly disclose wherein the retrieving a textual representation includes generating a request for information, sending the request to a server, and obtaining, from the server, at least the textual representation of the audio signal. However, *Foster* discloses a system for Interactive Machine Translation, where the user provides a translation of the source data, displayed as text, using a machine translation system as a resource. The use of the MT system suggests the use of a computer, including memory and a processor configured to execute instructions from memory. In addition, in any computer system software instructions, for example function calls, are executed in order to retrieve data from memory, such as a server, for further processing.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply the known technique of sending a request for information to a server and obtain a textual representation of the audio signal in *Foster*, since it would enable the user to process information previously stored in memory.

13. As per claims 3 and 23, **Foster** in view of **Schulz** disclose the method and system of claims 1 and 21, and **Schulz** further discloses wherein the presenting the textual representation to a user, includes: obtaining the audio signal, providing the audio signal and the textual representation of the audio signal to the user, and visually synchronizing the providing of the audio signal with the textual representation of the audio signal (column 5 lines 30-33 and column 6 lines 29-30, *the audio signal is provided the user, synchronized with the text. Therefore the audio signal must have first been obtained*). **Schulz** discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on a screen with a specific spoken word during playback of an audio file. All of the elements of claims 3 and 23 are known in the references **Foster** and **Schulz**, the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with **Foster**, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

14. As per claims 4 and 24, **Foster** in view of **Schulz** disclose the method and system of claims 3 and 23, and **Schulz** further discloses wherein the obtaining the audio signal includes

accessing a database of original media to retrieve the audio signal (column 5 lines 30-33, *the audio recording is played back and aligned with the words on the screen. The audio played back is from an audio recording; therefore the audio must have been accessed from a recording medium or memory, such as a database*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to access a database of original media to retrieve the audio signal in **Foster**, since it would enable the user to process information previously stored in the database.

15. As per claims 5,8,25, and 28 **Foster** in view of **Schulz** disclose the method and system of claims 3,1,23 and 21, and **Schulz** further discloses wherein the obtaining the audio signal includes receiving input, from the user, regarding a desire for the audio signal (column 12 line 63-column 13 line 12, *if the user enters a command to start playback of the audio signal, the playback edit function mode is entered, otherwise the system enters the standard editing mode*) initiating a media player, and using the media player to obtain the audio signal (column 12 line 63-column 13 line 12, *if the user enters a command to start playback of the audio signal, the playback edit function mode is entered and playback of the audio recording synchronized with the text begins. Since the audio, a type of media, is output, it must be have been obtained and output through a media player*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to receive input, from the user, regarding a desire for the audio signal, initialize a media player, and use the media player to obtain the audio signal in **Foster**, since it would enable

the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in *Schulz* (column 5 lines 55-58).

16. As per claims 6 and 26, *Foster* in view of *Schulz* disclose the method and system of claims 1 and 21, but *Foster* does not explicitly disclose wherein the receiving selection of a segment of the textual representation includes identifying a portion of the textual representation selected by the user, accessing a server to obtain text corresponding to the portion of the textual representation, and receiving, from the server, the text corresponding to the portion of the textual representation. However, *Foster* discloses a system for Interactive Machine Translation, where the user provides a translation of the source data using a machine translation system as a resource. The use of the MT system suggests the use of a computer, including memory and a processor configured to execute instructions from memory. In addition, in any computer system software instructions, for example function calls, are executed in order to retrieve data from memory, such as a server, for further processing.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply the known technique accessing and receiving text from a server in *Foster*, since it would enable the system to process information previously stored in memory.

17. As per claims 7 and 27, *Foster* in view of *Schulz* disclose the method and system of claims 6 and 26, and *Schulz* further discloses wherein the text includes a transcription of the

audio signal and metadata corresponding to the portion of the textual representation (column 4 lines 52-59, *a file containing the transcription of the input speech also contains beginning and end times for each word and silent pauses*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have text file that includes a transcription and metadata in **Foster**, since it would enable the system to locate pauses, and suppress them during playback, as indicated in **Schulz** (column 4 lines 60-65).

18. As per claims 9 and 29, **Foster** in view of **Schulz** disclose the method and system of claims 8 and 28, and **Schulz** further discloses wherein the using the media player includes identifying, by the media player, the segment of the textual representation, and retrieving the portion of the audio signal corresponding to the segment of the textual representation (column 6 lines 18-30, *the system uses the beginning and ending times of words to align the cursor on the monitor with a particular displayed word during playback of the audio recording. Since the audio is played back synchronized with the time information from the text file, a media player must have identified the textual representation and retrieved the audio signal*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to identify, by the media player, the segment of the textual representation, and retrieve the portion of the audio signal corresponding to the segment of the textual representation in **Foster**, since it would enable the user to quickly and easily translate and edit text displayed on

the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

19. As per claims 10,11, 30 and 31, **Foster** in view of **Schulz** disclose the method and system of claims 9 and 29, and **Schulz** further discloses wherein the segment of the textual representation includes a starting position in the textual representation, and wherein the identifying the segment includes identifying a time codes associated with the beginning and ending of the textual representation (column 6 lines 18-30, *the system uses the beginning and ending times of words to align the cursor on the monitor with a particular displayed word during playback of the audio recording*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a textual representation that includes a starting position, and identify time codes associated with the beginning and end times of the textual representation in **Foster**, since it would enable the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

20. As per claims 13 and 33, **Foster** in view of **Schulz** disclose the method and system of claims 1 and 21, and **Schulz** further discloses wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation

Art Unit: 2626

(column 5 lines 30-33 and column 6 lines 29-30). **Schulz** discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on the screen with a specific spoken word during playback of the audio file. All of the elements of claims 13 and 33 are known in the references **Foster** and **Schulz**, the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with **Foster**, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

21. As per claims 14 and 34, **Foster** in view of **Schulz** disclose the method and system of claims 13 and 33, and **Schulz** further discloses wherein the segment of the textual representation includes time codes corresponding to when words in the textual representation were spoken (column 4 lines 52-59, *a file containing the transcription of the input speech also contains beginning and end times for each word and silent pauses*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a textual representation that includes time codes corresponding to when words

in the textual representation were spoken in **Foster**, since it would enable the system to locate pauses, and suppress them during playback, as indicated in **Schulz** (column 4 lines 60-65).

22. As per claims 15 and 35, **Foster** in view of **Schulz** disclose the method and system of claims 14 and 34, and **Schulz** further discloses wherein the visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation includes comparing times corresponding to the providing of the portion of the audio signal to the time codes from the segment of the textual representation, and visually distinguishing words in the segment of the textual representation when the words are spoken during the providing of the portion of the audio signal (column 6 lines 18-30).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to compare times corresponding to the providing of the portion of the audio signal to the time codes from the segment of the textual representation, and visually distinguishing words in the segment of the textual representation when the words are spoken during the providing of the portion of the audio signal in **Foster**, since it would enable the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

23. As per claims 16,17,36 and 37, **Foster** in view of **Schulz** disclose the method of claims 1 and 21, and **Schulz** further discloses wherein the providing the segment of the textual

representation and the portion of the audio signal to the user includes permitting the user to control the providing of the portion of the audio signal by allowing the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal using foot pedals (column 2 lines 29-34).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to control the providing of the portion of the audio signal by allowing the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal using foot pedals in **Foster**, since it would enable the user to control playback of the audio file, thus and quickly and efficiently process the source data into target data.

24. As per claims 18 and 38, **Foster** in view of **Schulz** disclose the method of claims 16 and 36, and **Schulz** further discloses wherein the permitting the user to control the providing of the portion of the audio signal includes permitting the user to rewind the portion of the audio signal at least one of a predetermined amount of time and a predetermined amount of words (column 2 line 29-34, *the user can use keyboard input or a foot control to control the audio signal, including moving forward and rewinding*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to permit the user to rewind the portion of the audio signal at least one of a predetermined amount of time and a predetermined amount of words in **Foster**, since it would enable the user to control playback of the audio file, thus and quickly and efficiently process the source data into target data.

25. As per claim 40, **Foster** discloses a graphical user interface, comprising:

A text input section that includes text information in a first language (page 179, section 3, first paragraph, *the translator selects text, therefore a textual representation must have been input*);

A translation section that receives a translation actually made by the user into a second language (page 179, section 3, first paragraph, *the translator selects a portion of the source text, usually a sentence, and types in the translation*).

Foster does not disclose a transcription section that includes a transcription of non-text information in a first language, a translation section that receives a translation made by the user of the non-text information, and a play button that, when selected, causes the retrieval of the non-text information to be initiated, playing of the non-text information, and the playing of the non-text information to be visually synchronized with the transcription in the transcription section. However, speech recognition systems are commonly used to convert speech to text, as indicated in **Schulz** (column 1 lines 27-34, *speech recognition is used for transcription*). **Schulz** also discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on the screen with a specific spoken word during playback of the audio file. All of the elements of claim 40 are known in references **Foster** and **Schulz**, the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use known methods to retrieve a transcript of non-text information in a first language in *Foster*, since it would provide automatic transcription, saving transcription costs (Schulz, column 1 lines 27-34), while enabling a user to provide fast and accurate translation of speech data.

It would also have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with *Foster*, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in *Schulz* (column 5 lines 55-58).

26. As per claim 44, *Foster* in view of *Schulz* disclose the graphical user interface of claim 40, and *Schulz* further discloses wherein the play button further causes words in the transcription to be visually distinguished in synchronism with the words in the non-text information being played (column 6 lines 18-30).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a play button that causes words in the transcription to be visually distinguished in synchronism with the words in the non-text information being played in *Foster*, since it would enable the user to quickly and easily translate and edit text displayed on the monitor, including

Art Unit: 2626

identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in **Schulz** (column 5 lines 55-58).

27. As per claim 45, **Foster** in view of **Schulz** disclose the graphical user interface of claim 40, and **Schulz** further discloses wherein the non-text information includes at least one of audio and video (column 4 lines 46-59, *a speech recognition unit converts a recording of speech (audio non-text information) into a text file*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to process non-text information that includes at least one of audio and video in **Foster**, since it would enable the system to translate spoken language as well as textual documents.

Claims 41 and 46 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Foster** in view of **Schulz** as applied to claim 40 above, and further in view of **Saindon** (6,820,055).

28. **Foster** in view of **Schulz** disclose the graphical user interface of claim 40, however neither disclose wherein the transcription visually distinguishes names of people, places, and organizations and wherein the graphical user interface is associated with a word processing application. **Saindon** discloses a system for automated transcription and translation that processes text to visually distinguish the names of people, places and organizations using a word processor (column 16 lines 34-65, *the system processes the text to determine if all proper nouns are capitalized using software such as Microsoft word*). All of the elements of claims 41 and 46

are known in references *Foster*, *Schulz*, and *Saindon* the only difference is their combination for use in a translation system.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply the known technique of having a transcription that visually distinguishes names of people, places, and organizations and a graphical user interface is associated with a word processing application in *Foster* and *Schulz*, since it would enable the system to generate text that provides accurate translations, as indicated in *Saindon* (column 16 lines 38-40), using reliable commercially established software that is readily available.

Claims are 12, 19, 32, 39, 42, 43, and 47 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Foster* in view of *Schulz* as applied to claims 1, 21 and 40 above, and further in view of *Shiotani* (4,814,988).

29. As per claims 12 and 32, *Foster* in view of *Schulz* disclose the method and system of claims 1 and 21, however neither disclose wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes displaying the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal, including as a split screen in a translation window. *Shiotani* discloses wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes displaying the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal, including as a split screen in a translation window (column 2 lines 15-20 and Figure 4(a) and

4(b)). *Shiotani* discloses a machine translation system where the source string and target string appear side-by-side in the same window.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to display the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal, including as a split screen in a translation window in *Foster* and *Schulz*, since one of ordinary skill in the art has good reason to pursue the options within his or her technical grasp in order to achieve the predictable result of quickly and efficiently translating source information.

30. As per claims 19 and 39, *Foster* in view of *Schulz* disclose the method of claims 1 and 21, however neither explicitly disclose publishing the translation to a user-determined location. However, *Schulz* does disclose a text editor used to synchronize text and audio information when editing the textual information (column 5 lines 30-33). In text editing software, such as Microsoft word or open office, the user has many options once a document is complete. It can either be saved to a file, transmitted over the internet, printed on a screen, sent to a printer, or a combination thereof. In addition, *Shiotani* discloses sending the translation to a CRT display (user-defined location) (column 3 lines 2-4).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to publish the translation to a user-determined location in *Foster* and *Schulz*, since it would enable the user to save the translation for use at a later time, or output the translation from current use.

31. As per claims 42 and 43, *Foster* in view of *Schulz* disclose the graphical user interface of claim 40, but neither explicitly disclose a configuration button, that when selected, causes a window to be presented, the window permitting an amount of backup to be specified, the amount of backup including one of a predetermined amount of time and a predetermined number of words, and wherein the window further permits a name to be given for the translation and a location of publication to be specified. However, *Shiotani* does disclose a translation buffer for storing the result of translation of a selected portion of the input (column 2 lines 38-41). The translation buffer stores a predetermined number of words, i.e. the region of the text specified by the user and then translated. In addition, the use of a configuration button to present a window that permits a name to be given to a file and a location of publication to be specified is a feature of any text editing or word processing software, running on any of a number of operating systems, such as windows and Linux. The software enables the user to use the save button (configuration button), located under a file menu in a task bar, to choose a location in memory as well as a name for the file.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply the known technique of using a configuration button, that when selected, causes a window to be presented, the window permitting an amount of backup to be specified, the amount of backup including one of a predetermined amount of time and a predetermined number of words, and wherein the window further permits a name to be given for the translation and a location of publication to be specified in *Foster* and *Schulz*, since it would enable the

Art Unit: 2626

system to save the file in memory so that it can be easily retrieved for further processing in the future.

32. As per claim 47, **Foster** discloses a method comprising:

A user viewing a textual information in a first language (page 179, section 3, first paragraph, *the translator selects text in a first language to be translated*);

Said user actually translating said information thereby obtaining a translation in a second language (page 179, section 3, first paragraph, *the translator selects a portion of the source text, usually a sentence, and types in the translation*).

Foster does not disclose a user listening to an audio playback of information in a first language while viewing a textual transcription of said information in said first language on a transcription section of a graphical user interface (GUI), said textual transcription being synchronized with said audio playback, said user translating the audio playback of said information, said user using a different section of said graphical user interface (GUI) to display said translation while making said translation. However, speech recognition systems are commonly used to convert speech to text, as indicated in **Schulz** (column 1 lines 27-34, *speech recognition is used for transcription*). **Schulz** also discloses a system that synchronizes text with a specific spoken word during playback of an audio file (column 5 lines 30-33). In **Schulz**, a text editor is used that automatically aligns a cursor in the written text on the screen with a specific spoken word during playback of the audio file.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the known elements of audio and text synchronization with *Foster*, since the combination would produce the predictable result of enabling the user to quickly and easily translate and edit text displayed on the monitor, including identifying and correcting errors, without interruption during playback of the speech from an audio recording, as indicated in *Schulz* (column 5 lines 55-58).

Additionally, *Shiotani* discloses displaying the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal, including as a split screen in a translation window (column 2 lines 15-20 and Figure 4(a) and 4(b)). *Shiotani* discloses a machine translation system where the source string and target string appear side-by-side in the same window.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to display the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal, including as a split screen in a translation window in *Foster*, since one of ordinary skill in the art has good reason to pursue the options within his or her technical grasp in order to achieve the predictable result of quickly and efficiently translating source information.

Conclusion

10. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Matthew Baker whose telephone number is (571)270-1856. The examiner can normally be reached on 4-5-9, First Friday Off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Application/Control Number: 10/610,684
Art Unit: 2626

Page 28

/Matthew Baker/
Examiner, Art Unit 2626

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626